This Raw Listing contains the General Information Section and up to the first 5 pages.

ENTERED

SEQUENCE LISTING

(1) General Information

(i) APPLICANT: Yanagisawa, Masashi
Bergsma, Derk
Wilson, Shelagh
Brooks, David
Gellai, Miklos

(ii) TITLE OF THE INVENTION:NOVEL LIGANDS OF THE NEUROPEPTIDE RECEPTOR HFGAN72

(iii) NUMBER OF SEQUENCES:-21-

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: United States of America
(F) ZIP: 19406

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 08/938,548
(B) FILING DATE: 26-SEPT-1997
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/887,382
(B) FILING DATE: 2-JUL-1997

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/820,519
(B) FILING DATE: 19-MAR-1997

(A) APPLICATION NUMBER: 60/033,604
(B) FILING DATE: 17-DEC-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: King, William T.
(B) REGISTRATION NUMBER: 30,954
(C) REFERENCE/DOCKET NUMBER: ATG50037-2

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 610-270-5219
(B) TELEFAX: 610-270-4026
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1970 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AAAACATAAT GTGGGTCTCG CGTCTGCCTC TCTCCCGCCC CTAATTAGCA GCTGCCTCCC     60
TCCATATTGT CCCAGGCCAG CGCTTCTTTT GTGCTCCCAG ATTCCTGGGT GCAAGGTGGC    120
CTCATTAGTG CCCGGAGACC GCCCCATCTC CAGGGAGCAG ATAGACAGAC AAGGGGGTGA    180
TCAGGGGCAC AGTGATCCAA CCCTGGCCTC TGAACGCCGC AGCGGCCATT CCTTGGGCCC    240
AGCCTGGAGA CGGCCCCCCT GCAGCAGGCT AATCTTAGAC TTGCCTTTGT CTGGCCTGGG    300
TGTGGACGCA ATGTGCCTGT CAATTCCCCG CCACCTCAGA GCACTATAAA CCCCAGACCC    360
CTGGGAGTGG GTCACAATTG ACAGCCTCAA GGTTCCTGGC TTTTTGAACC ACCACAGACA    420
TCTCCTTTCC CGGCTACCCC ACCCTGAGCG CCAGACACCA TGAACCTTCC TTCCACAAAG    480
GTAAAGATCC AGGGATGGAG GGGTGACTCA GCCATCCCAG AGGAAGCAAA AAGAGTGCTT    540
GCTCAGAGGG CTGGAAGAAA GGCCAAAGGT GTCTCCACTC TTGGTCTTTT CCTGGGTGTG    600
CTCTGAGGCA GGAGCACCTG CCTTGGCTCA CATTGGGTTG GGTGCTGTTT TGCTAAGAGC    660
CTGTGTTTGC TGAGCTCATA TGTGTCAGGT GCTCCGTTTG CACCTGTCAT CTCTTGTCAT    720
CCTCCCAACA GCCTTGCAGA GTAGAAATTA TTTCTAGTAT ACCCAGTTTA CAGGTAAGGG    780
AGCTGTGCCC TCTGAAAGGG CAGGAAACTG GTTCAAAGCA ACGGAGTTCA GTCACTCCTG    840
CAAGGGGGCA GGCAGATGAG AGAGCATTCT GGAGTCTTGC TAGTTCCTGA TTTCCATGTG    900
TTTCCCTGCT GTGGAGAGGA AGTTGGGGGG ACTCAGTAGG GCCCGGGTTT TTCCCAAGTT    960
TACAACTTCT GCTGCAGACA GACACTCCTG TTTTCAGGTG GAGTGGCAAG TGCCCTAGTG   1020
GTGGCAACAG TGGCCTAAGT CTCCAGAGAA AAGGGGGATT CACTCTGCCC AGGGGGTCTC   1080
AAAAGGCTTC CTGTGGGAGA TGCTCTGCTG GGTCTTGAAG GAGGAGCAGG GAAAGTAGGC   1140
CGATACCAGC AAGGGCGCAA AGCAAGGAGA ACTAAGTGAC AGCCAGAAAG GAGTGCAGGC   1200
TTGGAGGGGG CGCGGAGCCA GAGGGGCAGG TCCTGTGCGT GGGAGCTGGT GGCGGGCGCC   1260
GTGGGAAGAC CCCCCCAGCG CCCTGTCTCC GTCTCCCTAG GTCTCCTGGG CCGCCGTGAC   1320
GCTACTGCTG CTGCTGCTGC TGCTGCCGCC CGCGCTGTTG TCGTCCGGGG CGGCTGCACA   1380
GCCCCTGCCC GACTGCTGTC GTCAAAAGAC TTGCTCTTGC CGCCTCTACG AGCTGCTGCA   1440
CGGCGCGGGC AATCACGCGG CCGGCATCCT CACGCTGGGC AAGCGGAGGT CCGGGCCCCC   1500
GGGCCTCCAG GGTCGGCTGC AGCGCCTCCT GCAGGCCAGC GGCAACCACG CCGCGGGCAT   1560
CCTGACCATG GGCCGCCGCG CAGGCGCAGA GCCAGCGCCG CGCCCCTGCC TCGGGCGCCG   1620
CTGTTCCGCC CCGGCCGCCG CCTCCGTCGC GCCCGGAGGA CAGTCCGGGA TCTGAGTCGT   1680
TCTTCGGGCC CTGTCCTGGC CCAGGCCTCT GCCCTCTGCC CACCCAGCGT CAGCCCCCAG   1740
AAAAAAGGCA ATAAAGACGA GTCTCCATTC GTGTGACTGG TCTCTGTTCC TGTGCGGTCG   1800
CGTCCTGCCC ATCCGGGGTG GCAAAGCGTC TTGCGGAGGA CAGCTGGGCC TGGAAGCCCG   1860
GCTGTCGGGC ACCAGCCTTA GCTTTTGCGT GGTTGAATCG GAAACACTCT TGGTTGGGGA   1920
```

GTTCCCAGTG CAAGGCCCTG GGGCACAGAG AGAACTGCAC AGGTGCATGC 1970

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 131 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Asn Leu Pro Ser Thr Lys Val Ser Trp Ala Ala Val Thr Leu Leu
 1               5                  10                  15
Leu Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Ser Gly Ala Ala
            20                  25                  30
Ala Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg
        35                  40                  45
Leu Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu
    50                  55                  60
Thr Leu Gly Lys Arg Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu
65                  70                  75                  80
Gln Arg Leu Leu Gln Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr
                85                  90                  95
Met Gly Arg Arg Ala Gly Ala Glu Pro Ala Pro Arg Pro Cys Leu Gly
            100                 105                 110
Arg Arg Cys Ser Ala Pro Ala Ala Ala Ser Val Ala Pro Gly Gly Gln
        115                 120                 125
Ser Gly Ile
    130
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
 1               5                  10                  15
Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
            20                  25                  30
Leu
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln Arg Leu Leu Gln
 1               5                  10                  15
Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr Met
            20                  25
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 585 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GGCTCGGCGG CCTCAGACTC CTTGGGTATT TGGACCACTG CACCGAAGAT ACCATCTCTC     60
CGGATTGCCT CTCCCTGAGC TCCAGACACC ATGAACCTTC CTTCTACAAA GGTTCCCTGG    120
GCCGCCGTGA CGCTGCTGCT GCTGCTACTG CTGCCGCCGG CGCTGCTGTC GCTTGGGGTG    180
GACGCGCAGC CTCTGCCCGA CTGCTGTCGC CAGAAGACGT GTTCCTGCCG TCTCTACGAA    240
CTGTTGCACG GAGCTGGCAA CCACGCCGCG GGCATCCTCA CTCTGGGAAA GCGGCGACCT    300
GGACCCCCAG GCCTCCAAGG ACGGCTGCAG CGCCTCCTTC AGGCCAACGG TAACCACGCA    360
GCTGGCATCC TGACCATGGG CCGCCGCGCA GGCGCAGAGC TAGAGCCATA TCCCTGCCCT    420
GGTCGCCGCT GTCCGACTGC AACCGCCACC GCTTTAGCGC CCCGGGGCGG ATCCAGAGTC    480
TGAACCCGTC TTCTATCCCT GTCCTAGTCC TAACTTTCCC CTCTCCTCGC CGGTCCCTAG    540
GCAATAAAGA CGTTTCTCTG CTAAAAAAAA AAAAAAAAAA AAAAA                    585
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
 1               5                  10                  15
Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
```

```
206                 20                  25                  30
207    Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
208             35                  40                  45
209    Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
210         50                  55                  60
211    Leu Gly Lys Arg Arg Pro Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln
212    65                  70                  75                  80
213    Arg Leu Leu Gln Ala Asn Gly Asn His Ala Ala Gly Ile Leu Thr Met
214                    85                  90                  95
215    Gly Arg Arg Ala Gly Ala Glu Leu Glu Pro Tyr Pro Cys Pro Gly Arg
216                100                 105                 110
217    Arg Cys Pro Thr Ala Thr Ala Thr Ala Leu Ala Pro Arg Gly Gly Ser
218            115                 120                 125
219    Arg Val
220        130
221
222              (2) INFORMATION FOR SEQ ID NO:7:
223
224           (i) SEQUENCE CHARACTERISTICS:
225             (A) LENGTH: 32 amino acids
226             (B) TYPE: amino acid
227             (C) STRANDEDNESS: single
228             (D) TOPOLOGY: linear
229
230           (ii) MOLECULE TYPE: protein
231
232           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
233
234    Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
235     1               5                  10                  15
236    Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
237                20                  25                  30
238
239              (2) INFORMATION FOR SEQ ID NO:8:
240
241           (i) SEQUENCE CHARACTERISTICS:
242             (A) LENGTH: 33 amino acids
243             (B) TYPE: amino acid
244             (C) STRANDEDNESS: single
245             (D) TOPOLOGY: linear
246
247           (ii) MOLECULE TYPE: protein
248
249           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
250
251    Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
252     1               5                  10                  15
253    Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
254                20                  25                  30
255    Leu
256
257
258              (2) INFORMATION FOR SEQ ID NO:9:
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION *US/08/938,548A*

DATE: 07/24/98
TIME: 12:44:17

INPUT SET: S27673.raw

Line	Error	Original Text